

# DNA SEQ SEARCH SUMMARY

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2004, 07:26:51 ; Search time 5556 Seconds  
(without alignments)  
10742.145 Million cell updates/sec

Title: US-09-955-315-1  
Perfect score: 1377  
Sequence: 1 ggggtgcgctctgcatcggtg.....aatcgctctaagggcctaa 1377

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*

Searched Seq 1 (DNA)  
2 (encoded PROT)

in DNA databases.

Also searched oligos

28: em\_un:\*  
 29: em\_vi:\*  
 30: em\_htg\_hum:\*  
 31: em\_htg\_inv:\*  
 32: em\_htg\_other:\*  
 33: em\_htg\_mus:\*  
 34: em\_htg\_pln:\*  
 35: em\_htg\_rod:\*  
 36: em\_htg\_mam:\*  
 37: em\_htg\_vrt:\*  
 38: em\_sy:\*  
 39: em\_htgo\_hum:\*  
 40: em\_htgo\_mus:\*  
 41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result |       | Query |        |        |    |           | Description         |
|--------|-------|-------|--------|--------|----|-----------|---------------------|
| No.    | Score | Match | Length | DB     | ID |           |                     |
|        | 1     | 1377  | 100.0  | 1377   | 6  | AX405298  | AX405298 Sequence   |
| c      | 2     | 1377  | 100.0  | 309400 | 6  | AX127153  | AX127153 Sequence   |
| c      | 3     | 1377  | 100.0  | 325651 | 1  | AP005283  | AP005283 Corynebact |
|        | 4     | 495   | 35.9   | 495    | 6  | AX123403  | AX123403 Sequence   |
|        | 5     | 495   | 35.9   | 495    | 6  | BD165520  | BD165520 Novel pol  |
| c      | 6     | 426.8 | 31.0   | 302070 | 1  | AP005223  | AP005223 Corynebact |
| c      | 7     | 358   | 26.0   | 813    | 6  | AX123404  | AX123404 Sequence   |
| c      | 8     | 358   | 26.0   | 813    | 6  | BD165521  | BD165521 Novel pol  |
| c      | 9     | 314.2 | 22.8   | 349659 | 1  | BX248360  | BX248360 Corynebact |
|        | 10    | 176.4 | 12.8   | 552    | 1  | AY065628  | AY065628 Mycobacte  |
|        | 11    | 169.8 | 12.3   | 4611   | 1  | AF068267  | AF068267 Streptomy  |
|        | 12    | 169.8 | 12.3   | 314100 | 1  | SCO939106 | AL939106 Streptomy  |
|        | 13    | 119.4 | 8.7    | 780    | 1  | H DU18769 | U18769 Haemophilus  |
| c      | 14    | 119.4 | 8.7    | 304558 | 1  | AE017151  | AE017151 Haemophil  |
|        | 15    | 118   | 8.6    | 10517  | 1  | AE014683  | AE014683 Bifidobact |
|        | 16    | 118   | 8.6    | 349980 | 6  | AX492783  | AX492783 Sequence   |
|        | 17    | 118   | 8.6    | 349980 | 6  | AX553950  | AX553950 Sequence   |
|        | 18    | 95.6  | 6.9    | 3657   | 1  | CGPTAACKA | X89084 C.glutamicu  |
| c      | 19    | 95.6  | 6.9    | 320550 | 1  | AP005282  | AP005282 Corynebact |
| c      | 20    | 95.6  | 6.9    | 349980 | 6  | AX127152  | AX127152 Sequence   |
|        | 21    | 86    | 6.2    | 309400 | 6  | AX127153  | AX127153 Sequence   |
|        | 22    | 86    | 6.2    | 325651 | 1  | AP005283  | AP005283 Corynebact |
|        | 23    | 86    | 6.2    | 349980 | 6  | AX127152  | AX127152 Sequence   |
| c      | 24    | 81.2  | 5.9    | 326150 | 1  | AP005281  | AP005281 Corynebact |
| c      | 25    | 81.2  | 5.9    | 349980 | 6  | AX127150  | AX127150 Sequence   |
|        | 26    | 74.2  | 5.4    | 531    | 6  | AR388806  | AR388806 Sequence   |
|        | 27    | 74.2  | 5.4    | 340000 | 1  | AP005274  | AP005274 Corynebact |
|        | 28    | 74.2  | 5.4    | 349980 | 6  | AX120085  | AX120085 Sequence   |
|        | 29    | 71    | 5.2    | 10146  | 1  | AE009194  | AE009194 Agrobacte  |
|        | 30    | 71    | 5.2    | 10161  | 1  | AE008160  | AE008160 Agrobacte  |
|        | 31    | 70.2  | 5.1    | 307050 | 1  | BX294140  | BX294140 Pirellula  |
|        | 32    | 70    | 5.1    | 504    | 1  | AY195736  | AY195736 Klebsiell  |
| c      | 33    | 68.4  | 5.0    | 299550 | 1  | AP005031  | AP005031 Streptomy  |

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2004, 02:55:31 ; Search time 593 Seconds  
(without alignments)  
9864.703 Million cell updates/sec

Title: US-09-955-315-1  
Perfect score: 1377  
Sequence: 1 ggggtgcgctctgcatcggtg.....aatgcctctaagggcctaa 1377

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_29Jan04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002s:\*  
7: geneseqn2003as:\*  
8: geneseqn2003bs:\*  
9: geneseqn2003cs:\*  
10: geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result |     |       | %     |        | Query |          | DB | ID | Description        |
|--------|-----|-------|-------|--------|-------|----------|----|----|--------------------|
|        | No. | Score | Match | Length | Match | Length   |    |    |                    |
|        | 1   | 1377  | 100.0 | 1377   | 6     | ABL56907 |    |    | Ab156907 Corynebac |
| c      | 2   | 1377  | 100.0 | 309400 | 5     | AAH68534 |    |    | Aah68534 C glutami |
|        | 3   | 495   | 35.9  | 495    | 5     | AAH68284 |    |    | Aah68284 C glutami |
|        | 4   | 495   | 35.9  | 495    | 7     | ACA00504 |    |    | Aca00504 C. glutam |
| c      | 5   | 441   | 32.0  | 441    | 7     | ACA00503 |    |    | Aca00503 C. glutam |
| c      | 6   | 358   | 26.0  | 813    | 5     | AAH68285 |    |    | Aah68285 C glutami |
| c      | 7   | 352   | 25.6  | 807    | 7     | ACA00505 |    |    | Aca00505 C. glutam |
|        | 8   | 273.6 | 19.9  | 483    | 7     | ACA29828 |    |    | Aca29828 Prokaryot |

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2004, 10:12:17 ; Search time 116 Seconds  
(without alignments)  
6587.650 Million cell updates/sec

Title: US-09-955-315-1  
Perfect score: 1377  
Sequence: 1 ggggtgcgctctgcacgcgtg.....aatcgctctaagggcctaa 1377

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | % Query |       | Match Length DB | ID | Description          |
|------------|---------|-------|-----------------|----|----------------------|
|            | Score   | Match |                 |    |                      |
| 1          | 74.2    | 5.4   | 531             | 4  | US-09-489-039A-5535  |
| 2          | 67      | 4.9   | 534             | 4  | US-09-543-681A-987   |
| 3          | 55.6    | 4.0   | 483             | 4  | US-09-252-991A-12631 |
| 4          | 45.4    | 3.3   | 501             | 4  | US-09-134-001C-1366  |
| 5          | 40.4    | 2.9   | 315             | 2  | US-08-889-013C-5     |
| c 6        | 39      | 2.8   | 1830121         | 4  | US-09-557-884-1      |
| c 7        | 39      | 2.8   | 1830121         | 4  | US-09-643-990A-1     |
| 8          | 38      | 2.8   | 564             | 4  | US-09-543-681A-150   |
| c 9        | 38      | 2.8   | 6436            | 4  | US-09-600-099-1      |
| 10         | 37      | 2.7   | 468             | 4  | US-09-252-991A-16447 |
| c 11       | 37      | 2.7   | 516             | 4  | US-09-252-991A-16153 |
| c 12       | 35.6    | 2.6   | 2889            | 4  | US-09-016-434-1192   |

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2004, 12:11:12 ; Search time 515 Seconds  
(without alignments)  
9362.649 Million cell updates/sec

Title: US-09-955-315-1  
Perfect score: 1377  
Sequence: 1 ggggtgcgctctgcacgcgtg.....aatcgctctaagggcctaa 1377

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2308684 seqs, 1750822206 residues

Total number of hits satisfying chosen parameters: 4617368

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_NA:\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
17: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
18: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result | No. | Score | %     |              | DB | ID                   | Description       |
|--------|-----|-------|-------|--------------|----|----------------------|-------------------|
|        |     |       | Query | Match Length |    |                      |                   |
| c      | 1   | 1377  | 100.0 | 1377         | 9  | US-09-955-315-1      | Sequence 1, Appli |
|        | 2   | 1377  | 100.0 | 3309400      | 9  | US-09-738-626-1      | Sequence 1, Appli |
|        | 3   | 495   | 35.9  | 495          | 9  | US-09-738-626-3319   | Sequence 3319, Ap |
| c      | 4   | 358   | 26.0  | 813          | 9  | US-09-738-626-3320   | Sequence 3320, Ap |
|        | 5   | 273.6 | 19.9  | 483          | 12 | US-10-282-122A-17698 | Sequence 17698, A |
|        | 6   | 176.4 | 12.8  | 546          | 12 | US-10-282-122A-25586 | Sequence 25586, A |
|        | 7   | 86    | 6.2   | 3309400      | 9  | US-09-738-626-1      | Sequence 1, Appli |
|        | 8   | 74.2  | 5.4   | 501          | 12 | US-10-282-122A-23746 | Sequence 23746, A |
|        | 9   | 68.4  | 5.0   | 468          | 14 | US-10-156-761-2491   | Sequence 2491, Ap |
| c      | 10  | 68.4  | 5.0   | 9025608      | 14 | US-10-156-761-1      | Sequence 1, Appli |
|        | 11  | 66.6  | 4.8   | 501          | 12 | US-10-282-122A-19300 | Sequence 19300, A |
|        | 12  | 66.4  | 4.8   | 504          | 12 | US-10-282-122A-41603 | Sequence 41603, A |
|        | 13  | 62    | 4.5   | 486          | 12 | US-10-282-122A-14089 | Sequence 14089, A |
|        | 14  | 61.8  | 4.5   | 501          | 12 | US-10-282-122A-36886 | Sequence 36886, A |
|        | 15  | 60.2  | 4.4   | 504          | 12 | US-10-282-122A-39666 | Sequence 39666, A |
| c      | 16  | 59.6  | 4.3   | 1364         | 9  | US-09-826-909-1      | Sequence 1, Appli |
|        | 17  | 58.8  | 4.3   | 504          | 12 | US-10-282-122A-20306 | Sequence 20306, A |
|        | 18  | 58.4  | 4.2   | 480          | 14 | US-10-156-761-1594   | Sequence 1594, Ap |
| c      | 19  | 58.4  | 4.2   | 9025608      | 14 | US-10-156-761-1      | Sequence 1, Appli |
|        | 20  | 57.8  | 4.2   | 177          | 9  | US-09-738-626-2911   | Sequence 2911, Ap |
|        | 21  | 55.6  | 4.0   | 471          | 9  | US-09-815-242-7717   | Sequence 7717, Ap |
|        | 22  | 55.6  | 4.0   | 471          | 12 | US-10-282-122A-30051 | Sequence 30051, A |
|        | 23  | 54.4  | 4.0   | 648          | 12 | US-10-282-122A-13139 | Sequence 13139, A |
|        | 24  | 52.8  | 3.8   | 486          | 12 | US-10-282-122A-11558 | Sequence 11558, A |
| c      | 25  | 51.2  | 3.7   | 807          | 14 | US-10-156-761-5411   | Sequence 5411, Ap |
|        | 26  | 45.4  | 3.3   | 444          | 12 | US-10-282-122A-34931 | Sequence 34931, A |
|        | 27  | 45.2  | 3.3   | 1377         | 9  | US-09-955-315-1      | Sequence 1, Appli |
|        | 28  | 44    | 3.2   | 436          | 12 | US-10-282-122A-10758 | Sequence 10758, A |
|        | 29  | 44    | 3.2   | 438          | 12 | US-10-282-122A-10356 | Sequence 10356, A |
|        | 30  | 42.8  | 3.1   | 558          | 12 | US-10-282-122A-15143 | Sequence 15143, A |
|        | 31  | 42.6  | 3.1   | 471          | 12 | US-10-282-122A-33816 | Sequence 33816, A |
|        | 32  | 40.4  | 2.9   | 2559         | 12 | US-10-282-122A-6960  | Sequence 6960, Ap |
|        | 33  | 40.2  | 2.9   | 471          | 12 | US-10-282-122A-31929 | Sequence 31929, A |
|        | 34  | 40.2  | 2.9   | 480          | 12 | US-10-282-122A-30920 | Sequence 30920, A |
|        | 35  | 39.6  | 2.9   | 174          | 9  | US-09-738-626-449    | Sequence 449, App |
|        | 36  | 39.2  | 2.8   | 415          | 12 | US-10-282-122A-22988 | Sequence 22988, A |
|        | 37  | 39    | 2.8   | 483          | 9  | US-09-815-242-7134   | Sequence 7134, Ap |
|        | 38  | 39    | 2.8   | 483          | 12 | US-10-282-122A-22313 | Sequence 22313, A |
| c      | 39  | 39    | 2.8   | 1830121      | 14 | US-10-329-960-1      | Sequence 1, Appli |
|        | 40  | 39    | 2.8   | 1830121      | 15 | US-10-329-670-1      | Sequence 1, Appli |
|        | 41  | 38.6  | 2.8   | 444          | 9  | US-09-815-242-8271   | Sequence 8271, Ap |
|        | 42  | 38.6  | 2.8   | 444          | 12 | US-10-282-122A-7708  | Sequence 7708, Ap |
|        | 43  | 38.4  | 2.8   | 932          | 12 | US-10-424-599-11756  | Sequence 11756, A |
|        | 44  | 38    | 2.8   | 552          | 12 | US-10-282-122A-33024 | Sequence 33024, A |
|        | 45  | 37.6  | 2.7   | 426          | 9  | US-09-815-242-4498   | Sequence 4498, Ap |

## ALIGNMENTS

RESULT 1  
US-09-955-315-1

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 21, 2004, 10:09:38 ; Search time 3900 Seconds  
(without alignments)  
10543.647 Million cell updates/sec

Title: US-09-955-315-1  
Perfect score: 1377  
Sequence: 1 ggggtgcgctctgcacgcgtg.....aatcgccctctaagggcctaa 1377

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gssl:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result |     |       | %     |        |       |          | DB | ID | Description        |
|--------|-----|-------|-------|--------|-------|----------|----|----|--------------------|
|        | No. | Score | Match | Length | Query |          |    |    |                    |
| c      | 1   | 59.4  | 4.3   | 636    | 14    | CF841965 |    |    | CF841965 psHB017xA |
| c      | 2   | 58.8  | 4.3   | 679    | 14    | CF851308 |    |    | CF851308 psMA019xF |
| c      | 3   | 58.8  | 4.3   | 693    | 14    | CF841926 |    |    | CF841926 psHB016xM |
| c      | 4   | 58.8  | 4.3   | 694    | 14    | CF851254 |    |    | CF851254 psMA019xB |
| c      | 5   | 58.8  | 4.3   | 705    | 14    | CF851365 |    |    | CF851365 psMA019xH |
| c      | 6   | 52    | 3.8   | 579    | 14    | CF754444 |    |    | CF754444 EST-77-2- |
| c      | 7   | 49.6  | 3.6   | 742    | 12    | BJ619993 |    |    | BJ619993 BJ619993  |
| c      | 8   | 45.6  | 3.3   | 350    | 9     | AU251612 |    |    | AU251612 AU251612  |
|        | 9   | 45.4  | 3.3   | 743    | 28    | BZ562850 |    |    | BZ562850 pacs2-164 |
| c      | 10  | 44.2  | 3.2   | 607    | 12    | BM415829 |    |    | BM415829 OP20911 M |
|        | 11  | 44    | 3.2   | 669    | 13    | CA092935 |    |    | CA092935 SCCCCL100 |
|        | 12  | 43    | 3.1   | 597    | 12    | BM415925 |    |    | BM415925 OP21007 M |
|        | 13  | 43    | 3.1   | 736    | 9     | AU168008 |    |    | AU168008 AU168008  |
|        | 14  | 42.8  | 3.1   | 842    | 28    | BZ571864 |    |    | BZ571864 msh2_2090 |
|        | 15  | 41.6  | 3.0   | 801    | 12    | BG825420 |    |    | BG825420 602746830 |
| c      | 16  | 41.4  | 3.0   | 506    | 12    | BJ191550 |    |    | BJ191550 BJ191550  |
| c      | 17  | 41.2  | 3.0   | 589    | 12    | BM415853 |    |    | BM415853 OP20935 M |
| c      | 18  | 40.4  | 2.9   | 927    | 28    | AZ692740 |    |    | AZ692740 ENTKY30TF |
| c      | 19  | 40.2  | 2.9   | 369    | 14    | W66612   |    |    | W66612 me23b10.r1  |
|        | 20  | 40    | 2.9   | 1168   | 13    | BX386385 |    |    | BX386385 BX386385  |
| c      | 21  | 39.2  | 2.8   | 566    | 6     | AL814237 |    |    | AL814237 Triticum  |
|        | 22  | 39    | 2.8   | 1201   | 13    | BX381961 |    |    | BX381961 BX381961  |
|        | 23  | 38.6  | 2.8   | 765    | 12    | BG340063 |    |    | BG340063 602438227 |
| c      | 24  | 38.6  | 2.8   | 1284   | 14    | CA207626 |    |    | CA207626 SCEQSB1C1 |
|        | 25  | 38.4  | 2.8   | 454    | 13    | BQ252468 |    |    | BQ252468 san76a09. |
|        | 26  | 38.4  | 2.8   | 482    | 13    | BQ295817 |    |    | BQ295817 sao27c02. |
|        | 27  | 38.4  | 2.8   | 525    | 12    | BM094323 |    |    | BM094323 saj14a09. |
|        | 28  | 38.4  | 2.8   | 844    | 28    | BZ571748 |    |    | BZ571748 msh2_2038 |
| c      | 29  | 38.2  | 2.8   | 388    | 10    | AW661443 |    |    | AW661443 833006B05 |
| c      | 30  | 38.2  | 2.8   | 664    | 28    | BZ563917 |    |    | BZ563917 pacs2-164 |
|        | 31  | 38.2  | 2.8   | 682    | 12    | BG450547 |    |    | BG450547 NF031A06D |
| c      | 32  | 38.2  | 2.8   | 915    | 29    | CG901851 |    |    | CG901851 ZMMBBb051 |
|        | 33  | 38.2  | 2.8   | 3137   | 11    | BC009370 |    |    | BC009370 Homo sapi |
| c      | 34  | 37.6  | 2.7   | 985    | 29    | CNS03SP8 |    |    | AL258821 Tetraodon |
| c      | 35  | 37.6  | 2.7   | 989    | 13    | BU345760 |    |    | BU345760 603524814 |
| c      | 36  | 37.6  | 2.7   | 1159   | 29    | CNS015XR |    |    | AL106041 Drosophil |
| c      | 37  | 37.2  | 2.7   | 885    | 13    | BX425603 |    |    | BX425603 BX425603  |
|        | 38  | 37    | 2.7   | 510    | 14    | CF841740 |    |    | CF841740 psHB015xM |
|        | 39  | 37    | 2.7   | 562    | 13    | BQ801298 |    |    | BQ801298 WHE2812_G |
|        | 40  | 37    | 2.7   | 690    | 29    | CNS0040D |    |    | AL054318 Drosophil |
| c      | 41  | 37    | 2.7   | 692    | 29    | CC940028 |    |    | CC940028 ZMMBBb023 |
| c      | 42  | 37    | 2.7   | 877    | 28    | BZ177450 |    |    | BZ177450 CH230-443 |
|        | 43  | 37    | 2.7   | 982    | 28    | CC452499 |    |    | CC452499 ZMMBBb034 |
|        | 44  | 36.8  | 2.7   | 450    | 14    | CA666218 |    |    | CA666218 wlk1.pk00 |
| c      | 45  | 36.8  | 2.7   | 465    | 9     | AA835142 |    |    | AA835142 ak63g06.s |



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 21, 2004, 19:17:56 ; Search time 3496 Seconds  
(without alignments)  
2045.652 Million cell updates/sec

Title: US-09-955-315-2  
Perfect score: 865  
Sequence: 1 MANYTVPGINENDAKQLIDG.....EKFQWFIRAHIVDVGNIQE 165

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delect 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+\_p2n.model -DEV=xlh

-

Q=/cgn2\_1/USPTO\_spool/US09955315/runat\_19022004\_145901\_28329/app\_query.fasta\_1.3  
27

-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09955315\_@CGN\_1\_1\_2496\_@runat\_19022004\_145901\_28329 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*

```

11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result |     |       | %     |        | Query |           | ID       | Description |
|--------|-----|-------|-------|--------|-------|-----------|----------|-------------|
|        | No. | Score | Match | Length | DB    |           |          |             |
|        | 1   | 865   | 100.0 | 495    | 6     | AX123403  | AX123403 | Sequence    |
|        | 2   | 865   | 100.0 | 495    | 6     | BD165520  | BD165520 | Novel pol   |
|        | 3   | 865   | 100.0 | 1377   | 6     | AX405298  | AX405298 | Sequence    |
| c      | 4   | 865   | 100.0 | 309400 | 6     | AX127153  | AX127153 | Sequence    |
| c      | 5   | 865   | 100.0 | 325651 | 1     | AP005283  | AP005283 | Corynebac   |
| c      | 6   | 690   | 79.8  | 302070 | 1     | AP005223  | AP005223 | Corynebac   |
| c      | 7   | 657   | 76.0  | 349659 | 1     | BX248360  | BX248360 | Corynebac   |
|        | 8   | 470   | 54.3  | 552    | 1     | AY065628  | AY065628 | Mycobacte   |
|        | 9   | 432   | 49.9  | 4611   | 1     | AF068267  | AF068267 | Streptomy   |
|        | 10  | 432   | 49.9  | 314100 | 1     | SCO939106 | AL939106 | Streptomy   |
|        | 11  | 364   | 42.1  | 780    | 1     | H DU18769 | U18769   | Haemophilus |
| c      | 12  | 364   | 42.1  | 304558 | 1     | AE017151  | AE017151 | Haemophil   |
|        | 13  | 328.5 | 38.0  | 10517  | 1     | AE014683  | AE014683 | Bifidobac   |
|        | 14  | 328.5 | 38.0  | 349980 | 6     | AX492783  | AX492783 | Sequence    |
|        | 15  | 328.5 | 38.0  | 349980 | 6     | AX553950  | AX553950 | Sequence    |
|        | 16  | 264   | 30.5  | 534    | 6     | AR375981  | AR375981 | Sequence    |

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 21, 2004, 19:17:25 ; Search time 370 Seconds  
(without alignments)  
1894.467 Million cell updates/sec

Title: US-09-955-315-2  
Perfect score: 865  
Sequence: 1 MANYTVPGINENDAKQLIDG.....EKFQWFIRAHIVDVGNIQE 165

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+\_p2n.model -DEV=xlh

Q=/cgn2\_1/USPTO\_spool/US09955315/runat\_19022004\_145901\_28319/app\_query.fasta\_1.3  
27

-DB=N\_Geneseq\_29Jan04 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09955315\_@CGN\_1\_1\_470\_@runat\_19022004\_145901\_28319 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_29Jan04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002s:\*  
7: geneseqn2003as:\*  
8: geneseqn2003bs:\*  
9: geneseqn2003cs:\*  
10: geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result | No. | Score | %<br>Query<br>Match | Length | DB | ID          | Description        |
|--------|-----|-------|---------------------|--------|----|-------------|--------------------|
|        | 1   | 865   | 100.0               | 495    | 5  | AAH68284    | Aah68284 C glutami |
|        | 2   | 865   | 100.0               | 495    | 7  | ACA00504    | Aca00504 C. glutam |
|        | 3   | 865   | 100.0               | 1377   | 6  | ABL56907    | Abl56907 Corynebac |
| c      | 4   | 865   | 100.0               | 309400 | 5  | AAH68534    | Aah68534 C glutami |
|        | 5   | 657   | 76.0                | 483    | 7  | ACA29828    | Aca29828 Prokaryot |
|        | 6   | 459   | 53.1                | 546    | 7  | ACA37716    | Aca37716 Prokaryot |
|        | 7   | 328.5 | 38.0                | 349980 | 6  | ABQ81845    | Abq81845 Bifidobac |
|        | 8   | 261   | 30.2                | 501    | 7  | ACA49016    | Aca49016 Prokaryot |
|        | 9   | 261   | 30.2                | 504    | 7  | ACA51796    | Aca51796 Prokaryot |
|        | 10  | 247   | 28.6                | 501    | 7  | ACA35876    | Aca35876 Prokaryot |
|        | 11  | 246   | 28.4                | 504    | 7  | ACA32436    | Aca32436 Prokaryot |
|        | 12  | 245   | 28.3                | 501    | 7  | ACA31430    | Aca31430 Prokaryot |
|        | 13  | 240   | 27.7                | 1206   | 5  | AAS82058    | Aas82058 DNA encod |
|        | 14  | 233   | 26.9                | 504    | 7  | ACA53733    | Aca53733 Prokaryot |
| c      | 15  | 204   | 23.6                | 441    | 7  | ACA00503    | Aca00503 C. glutam |
|        | 16  | 199   | 23.0                | 29139  | 4  | AAS59569    | Aas59569 Propionib |
|        | 17  | 199   | 23.0                | 29139  | 7  | ACF64498    | Acf64498 Propionib |
|        | 18  | 182   | 21.0                | 438    | 7  | ACA47774    | Aca47774 Prokaryot |
|        | 19  | 180   | 20.8                | 110000 | 7  | ACF67367_16 | Continuation (17 o |
|        | 20  | 180   | 20.8                | 210710 | 7  | ACF65380    | Acf65380 Photorhab |
|        | 21  | 179   | 20.7                | 438    | 7  | ACA22486    | Aca22486 Prokaryot |
|        | 22  | 175   | 20.2                | 415    | 7  | ACA35118    | Aca35118 Prokaryot |
|        | 23  | 169   | 19.5                | 444    | 7  | ACA47061    | Aca47061 Prokaryot |
|        | 24  | 169   | 19.5                | 486    | 7  | ACA23688    | Aca23688 Prokaryot |
| c      | 25  | 169   | 19.5                | 2962   | 4  | AAH54147    | Aah54147 S. epider |
|        | 26  | 168   | 19.4                | 471    | 7  | ACA44059    | Aca44059 Prokaryot |
|        | 27  | 167.5 | 19.4                | 444    | 4  | AAS54634    | Aas54634 Staphyloc |
|        | 28  | 166   | 19.2                | 426    | 4  | AAS51916    | Aas51916 Staphyloc |
|        | 29  | 165.5 | 19.1                | 441    | 7  | ACF73436    | Acf73436 Staphyloc |
|        | 30  | 165.5 | 19.1                | 444    | 7  | ACA19838    | Aca19838 Prokaryot |
|        | 31  | 163   | 18.8                | 501    | 6  | ABN91903    | Abn91903 Staphyloc |
| c      | 32  | 161   | 18.6                | 110000 | 6  | ABA90521_20 | Continuation (21 o |
| c      | 33  | 161   | 18.6                | 110000 | 6  | ABA90521_21 | Continuation (22 o |
|        | 34  | 160   | 18.5                | 525    | 7  | ACA47978    | Aca47978 Prokaryot |
|        | 35  | 160   | 18.5                | 660    | 6  | AAL41884    | Aal41884 Streptoco |
| c      | 36  | 159.5 | 18.4                | 17075  | 6  | ABL56780    | Abl56780 Nucleotid |
|        | 37  | 159   | 18.4                | 471    | 7  | ACA45946    | Aca45946 Prokaryot |
|        | 38  | 159   | 18.4                | 480    | 9  | ADC91885    | Adc91885 E. faeciu |
|        | 39  | 157   | 18.2                | 486    | 7  | ACA26219    | Aca26219 Prokaryot |
|        | 40  | 156   | 18.0                | 648    | 7  | ACA25269    | Aca25269 Prokaryot |
|        | 41  | 155   | 17.9                | 348    | 7  | ACF68866    | Acf68866 Photorhab |
|        | 42  | 155   | 17.9                | 465    | 7  | ACA32822    | Aca32822 Prokaryot |
|        | 43  | 155   | 17.9                | 525    | 6  | ABN70095    | Abn70095 Streptoco |
|        | 44  | 155   | 17.9                | 528    | 7  | ACA50604    | Aca50604 Prokaryot |
|        | 45  | 152   | 17.6                | 436    | 7  | ACA22888    | Aca22888 Prokaryot |

# ALIGNMENTS

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 21, 2004, 19:21:16 ; Search time 75 Seconds  
(without alignments)  
1220.892 Million cell updates/sec

Title: US-09-955-315-2  
Perfect score: 865  
Sequence: 1 MANYTVPGINENDAKQLIDG.....EKFQWFIRAHIVDVGNIQE 165

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+\_p2n.model -DEV=xlh

Q=/cgn2\_1/USPTO\_spool/US09955315/runat\_19022004\_145902\_28366/app\_query.fasta\_1.3  
27

-DB=Issued\_Patents\_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09955315\_@CGN\_1\_1\_56\_@runat\_19022004\_145902\_28366 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

| Result |       | %           |         |    |                      |                   |  |
|--------|-------|-------------|---------|----|----------------------|-------------------|--|
| No.    | Score | Query Match | Length  | DB | ID                   | Description       |  |
| 1      | 264   | 30.5        | 534     | 4  | US-09-543-681A-987   | Sequence 987, App |  |
| 2      | 247   | 28.6        | 531     | 4  | US-09-489-039A-5535  | Sequence 5535, Ap |  |
| 3      | 163   | 18.8        | 501     | 4  | US-09-134-001C-1366  | Sequence 1366, Ap |  |
| c 4    | 160   | 18.5        | 5301    | 4  | US-08-956-171E-443   | Sequence 443, App |  |
| 5      | 159   | 18.4        | 480     | 4  | US-09-107-532A-1512  | Sequence 1512, Ap |  |
| 6      | 151   | 17.5        | 483     | 4  | US-09-252-991A-12631 | Sequence 12631, A |  |
| c 7    | 142.5 | 16.5        | 21040   | 4  | US-08-961-527-55     | Sequence 55, Appl |  |
| c 8    | 134.5 | 15.5        | 1830121 | 4  | US-09-557-884-1      | Sequence 1, Appli |  |
| c 9    | 134.5 | 15.5        | 1830121 | 4  | US-09-643-990A-1     | Sequence 1, Appli |  |
| 10     | 133.5 | 15.4        | 315     | 2  | US-08-889-013C-5     | Sequence 5, Appli |  |
| 11     | 122.5 | 14.2        | 564     | 4  | US-09-543-681A-150   | Sequence 150, App |  |
| 12     | 114.5 | 13.2        | 576     | 4  | US-09-107-532A-1728  | Sequence 1728, Ap |  |
| 13     | 85    | 9.8         | 3048    | 1  | US-08-188-228-47     | Sequence 47, Appl |  |
| 14     | 85    | 9.8         | 3048    | 1  | US-08-332-643-41     | Sequence 41, Appl |  |
| 15     | 85    | 9.8         | 3048    | 1  | US-08-332-638-47     | Sequence 47, Appl |  |
| 16     | 83    | 9.6         | 1845    | 4  | US-09-543-681A-2894  | Sequence 2894, Ap |  |
| c 17   | 83    | 9.6         | 4403765 | 3  | US-09-103-840A-2     | Sequence 2, Appli |  |
| c 18   | 83    | 9.6         | 4411529 | 3  | US-09-103-840A-1     | Sequence 1, Appli |  |
| 19     | 82    | 9.5         | 1422    | 4  | US-09-252-991A-10517 | Sequence 10517, A |  |
| c 20   | 81.5  | 9.4         | 1208    | 4  | US-09-205-258-25     | Sequence 25, Appl |  |
| 21     | 81    | 9.4         | 2022    | 4  | US-09-328-352-1708   | Sequence 1708, Ap |  |
| 22     | 79.5  | 9.2         | 1203    | 4  | US-09-134-001C-1483  | Sequence 1483, Ap |  |
| 23     | 79    | 9.1         | 1664976 | 4  | US-08-916-421B-1     | Sequence 1, Appli |  |
| 24     | 78    | 9.0         | 1020    | 4  | US-09-107-532A-1490  | Sequence 1490, Ap |  |
| c 25   | 77    | 8.9         | 957     | 4  | US-09-369-247-56     | Sequence 56, Appl |  |
| 26     | 77    | 8.9         | 1476    | 1  | US-08-489-733-4      | Sequence 4, Appli |  |
| 27     | 77    | 8.9         | 1476    | 2  | US-08-993-581B-4     | Sequence 4, Appli |  |
| 28     | 77    | 8.9         | 1727    | 3  | US-08-999-733-2      | Sequence 2, Appli |  |
| 29     | 77    | 8.9         | 9757    | 1  | US-08-093-453B-1     | Sequence 1, Appli |  |
| 30     | 77    | 8.9         | 9759    | 1  | US-08-459-041A-1     | Sequence 1, Appli |  |
| 31     | 77    | 8.9         | 9759    | 3  | US-08-999-733-1      | Sequence 1, Appli |  |
| 32     | 76    | 8.8         | 1554    | 4  | US-09-134-001C-2607  | Sequence 2607, Ap |  |
| 33     | 75.5  | 8.7         | 1398    | 4  | US-09-489-039A-1374  | Sequence 1374, Ap |  |
| 34     | 75.5  | 8.7         | 1932    | 4  | US-09-252-991A-1911  | Sequence 1911, Ap |  |
| 35     | 75.5  | 8.7         | 2505    | 4  | US-09-252-991A-1830  | Sequence 1830, Ap |  |
| c 36   | 75    | 8.7         | 642     | 4  | US-09-370-838-119    | Sequence 119, App |  |
| c 37   | 75    | 8.7         | 808     | 4  | US-09-023-655-1279   | Sequence 1279, Ap |  |
| 38     | 75    | 8.7         | 2368    | 4  | US-09-343-011B-3     | Sequence 3, Appli |  |
| 39     | 75    | 8.7         | 4539    | 4  | US-09-328-352-1037   | Sequence 1037, Ap |  |
| 40     | 74.5  | 8.6         | 873     | 4  | US-09-252-991A-2784  | Sequence 2784, Ap |  |
| 41     | 74.5  | 8.6         | 1468    | 2  | US-09-074-512-2      | Sequence 2, Appli |  |
| c 42   | 74.5  | 8.6         | 1536    | 4  | US-09-252-991A-3182  | Sequence 3182, Ap |  |
| 43     | 74.5  | 8.6         | 2070    | 4  | US-09-252-991A-2687  | Sequence 2687, Ap |  |
| 44     | 74.5  | 8.6         | 4403765 | 3  | US-09-103-840A-2     | Sequence 2, Appli |  |
| 45     | 74.5  | 8.6         | 4411529 | 3  | US-09-103-840A-1     | Sequence 1, Appli |  |

## ALIGNMENTS

RESULT 1

US-09-543-681A-987

; Sequence 987, Application US/09543681A

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 21, 2004, 20:33:27 ; Search time 328 Seconds  
(without alignments)  
1761.498 Million cell updates/sec

Title: US-09-955-315-2  
Perfect score: 865  
Sequence: 1 MANYTVPGINENDAKQLIDG.....EKFQWFIRAHIVDVDGNIQE 165

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delect 7.0

Searched: 2308684 seqs, 1750822206 residues

Total number of hits satisfying chosen parameters: 4617368

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+\_p2n.model -DEV=xlh

Q=/cgn2\_1/USPTO\_spool/US09955315/runat\_19022004\_145904\_28498/app\_query.fasta\_1.3  
27

-DB=Published\_Applications\_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09955315\_@CGN\_1\_1\_57\_@runat\_19022004\_145904\_28498  
-NCPU=6 -ICPU=3 -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published\_Applications\_NA:\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*

12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
 13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*  
 14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*  
 15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
 16: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
 17: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
 18: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result |     |       | %     |         |    |                      | Description       |
|--------|-----|-------|-------|---------|----|----------------------|-------------------|
|        | No. | Score | Match | Length  | DB | ID                   |                   |
| c      | 1   | 865   | 100.0 | 495     | 9  | US-09-738-626-3319   | Sequence 3319, Ap |
|        | 2   | 865   | 100.0 | 1377    | 9  | US-09-955-315-1      | Sequence 1, Appli |
|        | 3   | 865   | 100.0 | 3309400 | 9  | US-09-738-626-1      | Sequence 1, Appli |
|        | 4   | 657   | 76.0  | 483     | 12 | US-10-282-122A-17698 | Sequence 17698, A |
|        | 5   | 459   | 53.1  | 546     | 12 | US-10-282-122A-25586 | Sequence 25586, A |
|        | 6   | 261   | 30.2  | 501     | 12 | US-10-282-122A-36886 | Sequence 36886, A |
|        | 7   | 261   | 30.2  | 504     | 12 | US-10-282-122A-39666 | Sequence 39666, A |
|        | 8   | 247   | 28.6  | 501     | 12 | US-10-282-122A-23746 | Sequence 23746, A |
|        | 9   | 246   | 28.4  | 504     | 12 | US-10-282-122A-20306 | Sequence 20306, A |
|        | 10  | 245   | 28.3  | 501     | 12 | US-10-282-122A-19300 | Sequence 19300, A |
| c      | 11  | 233   | 26.9  | 504     | 12 | US-10-282-122A-41603 | Sequence 41603, A |
|        | 12  | 231   | 26.7  | 468     | 14 | US-10-156-761-2491   | Sequence 2491, Ap |
|        | 13  | 231   | 26.7  | 9025608 | 14 | US-10-156-761-1      | Sequence 1, Appli |
|        | 14  | 227.5 | 26.3  | 480     | 14 | US-10-156-761-1594   | Sequence 1594, Ap |
|        | 15  | 227.5 | 26.3  | 9025608 | 14 | US-10-156-761-1      | Sequence 1, Appli |
|        | 16  | 182   | 21.0  | 438     | 12 | US-10-282-122A-35644 | Sequence 35644, A |
|        | 17  | 179   | 20.7  | 438     | 12 | US-10-282-122A-10356 | Sequence 10356, A |
|        | 18  | 175   | 20.2  | 415     | 12 | US-10-282-122A-22988 | Sequence 22988, A |
|        | 19  | 169   | 19.5  | 444     | 12 | US-10-282-122A-34931 | Sequence 34931, A |
|        | 20  | 169   | 19.5  | 486     | 12 | US-10-282-122A-11558 | Sequence 11558, A |
| c      | 21  | 168   | 19.4  | 471     | 12 | US-10-282-122A-31929 | Sequence 31929, A |
|        | 22  | 167.5 | 19.4  | 444     | 9  | US-09-815-242-8271   | Sequence 8271, Ap |
|        | 23  | 166   | 19.2  | 426     | 9  | US-09-815-242-4498   | Sequence 4498, Ap |
|        | 24  | 165.5 | 19.1  | 444     | 12 | US-10-282-122A-7708  | Sequence 7708, Ap |
|        | 25  | 160   | 18.5  | 525     | 12 | US-10-282-122A-35848 | Sequence 35848, A |
|        | 26  | 160   | 18.5  | 5301    | 8  | US-08-781-986A-443   | Sequence 443, App |
|        | 27  | 159   | 18.4  | 471     | 12 | US-10-282-122A-33816 | Sequence 33816, A |
|        | 28  | 157   | 18.2  | 486     | 12 | US-10-282-122A-14089 | Sequence 14089, A |
|        | 29  | 156   | 18.0  | 648     | 12 | US-10-282-122A-13139 | Sequence 13139, A |
|        | 30  | 155   | 17.9  | 465     | 12 | US-10-282-122A-20692 | Sequence 20692, A |
| c      | 31  | 155   | 17.9  | 528     | 12 | US-10-282-122A-38474 | Sequence 38474, A |
|        | 32  | 152   | 17.6  | 436     | 12 | US-10-282-122A-10758 | Sequence 10758, A |
|        | 33  | 151   | 17.5  | 471     | 9  | US-09-815-242-7717   | Sequence 7717, Ap |
|        | 34  | 151   | 17.5  | 471     | 12 | US-10-282-122A-30051 | Sequence 30051, A |
|        | 35  | 150   | 17.3  | 465     | 12 | US-10-282-122A-20841 | Sequence 20841, A |
|        | 36  | 148   | 17.1  | 537     | 9  | US-09-815-242-9396   | Sequence 9396, Ap |
|        | 37  | 148   | 17.1  | 537     | 12 | US-10-282-122A-37912 | Sequence 37912, A |
|        | 38  | 146   | 16.9  | 439     | 9  | US-09-974-300-6060   | Sequence 6060, Ap |
|        | 39  | 142.5 | 16.5  | 21040   | 12 | US-10-158-844-55     | Sequence 55, Appl |
|        | 40  | 138   | 16.0  | 438     | 9  | US-09-974-300-1602   | Sequence 1602, Ap |



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 21, 2004, 19:21:01 ; Search time 2365 Seconds  
(without alignments)  
2083.408 Million cell updates/sec

Title: US-09-955-315-2  
Perfect score: 865  
Sequence: 1 MANYTVPGINENDAKQLIDG.....EKFQWFIRAHIVDVDGNIQE 165

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delect 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+\_p2n.model -DEV=xlh

Q=/cgn2\_1/USPTO\_spool/US09955315/runat\_19022004\_145902\_28339/app\_query.fasta\_1.3  
27

-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09955315\_@CGN\_1\_1\_1906\_@runat\_19022004\_145902\_28339 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*

12: gb\_est3:\*  
 13: gb\_est4:\*  
 14: gb\_est5:\*  
 15: em\_estfun:\*  
 16: em\_estom:\*  
 17: em\_gss\_hum:\*  
 18: em\_gss\_inv:\*  
 19: em\_gss\_pln:\*  
 20: em\_gss\_vrt:\*  
 21: em\_gss\_fun:\*  
 22: em\_gss\_mam:\*  
 23: em\_gss\_mus:\*  
 24: em\_gss\_pro:\*  
 25: em\_gss\_rod:\*  
 26: em\_gss\_phg:\*  
 27: em\_gss\_vrl:\*  
 28: gb\_gss1:\*  
 29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

|        |     | %     |       | Query  |    |          | DB | ID                 | Description |
|--------|-----|-------|-------|--------|----|----------|----|--------------------|-------------|
| Result | No. | Score | Match | Length |    |          |    |                    |             |
| c      | 1   | 246   | 28.4  | 636    | 14 | CF841965 |    | CF841965 psHB017xA |             |
| c      | 2   | 246   | 28.4  | 679    | 14 | CF851308 |    | CF851308 psMA019xF |             |
| c      | 3   | 246   | 28.4  | 693    | 14 | CF841926 |    | CF841926 psHB016xM |             |
| c      | 4   | 246   | 28.4  | 694    | 14 | CF851254 |    | CF851254 psMA019xB |             |
| c      | 5   | 246   | 28.4  | 705    | 14 | CF851365 |    | CF851365 psMA019xH |             |
| c      | 6   | 221   | 25.5  | 742    | 12 | BJ619993 |    | BJ619993 BJ619993  |             |
| c      | 7   | 189   | 21.8  | 607    | 12 | BM415829 |    | BM415829 OP20911 M |             |
| c      | 8   | 186   | 21.5  | 350    | 9  | AU251612 |    | AU251612 AU251612  |             |
|        | 9   | 174   | 20.1  | 597    | 12 | BM415925 |    | BM415925 OP21007 M |             |
|        | 10  | 168.5 | 19.5  | 669    | 13 | CA092935 |    | CA092935 SCCCCL100 |             |
| c      | 11  | 166   | 19.2  | 589    | 12 | BM415853 |    | BM415853 OP20935 M |             |
|        | 12  | 148   | 17.1  | 3137   | 11 | BC009370 |    | BC009370 Homo sapi |             |
|        | 13  | 144   | 16.6  | 736    | 9  | AU168008 |    | AU168008 AU168008  |             |
| c      | 14  | 131   | 15.1  | 464    | 14 | CF851326 |    | CF851326 psMA019xF |             |
|        | 15  | 121   | 14.0  | 743    | 28 | BZ562850 |    | BZ562850 pacs2-164 |             |
|        | 16  | 118   | 13.6  | 897    | 28 | BH770759 |    | BH770759 LLMGtag50 |             |
|        | 17  | 102.5 | 11.8  | 1116   | 12 | BM415949 |    | BM415949 OP21031 M |             |
| c      | 18  | 96    | 11.1  | 676    | 14 | CB854133 |    | CB854133 UI-CF-DU1 |             |
| c      | 19  | 96    | 11.1  | 688    | 14 | CF891147 |    | CF891147 UI-CF-DU1 |             |
|        | 20  | 95.5  | 11.0  | 842    | 28 | BZ571864 |    | BZ571864 msh2_2090 |             |
| c      | 21  | 92    | 10.6  | 369    | 14 | W66612   |    | W66612 me23b10.r1  |             |
|        | 22  | 91.5  | 10.6  | 844    | 28 | BZ571748 |    | BZ571748 msh2_2038 |             |
|        | 23  | 90.5  | 10.5  | 3526   | 11 | AK054172 |    | AK054172 Mus muscu |             |
|        | 24  | 90.5  | 10.5  | 4982   | 11 | AK046971 |    | AK046971 Mus muscu |             |
|        | 25  | 90    | 10.4  | 310    | 10 | BE711894 |    | BE711894 QV2-HT069 |             |
|        | 26  | 87.5  | 10.1  | 675    | 12 | BI109593 |    | BI109593 602897552 |             |
| c      | 27  | 87.5  | 10.1  | 1019   | 9  | AL530716 |    | AL530716 AL530716  |             |
|        | 28  | 87    | 10.1  | 599    | 28 | BZ304474 |    | BZ304474 KD2968.pl |             |
|        | 29  | 87    | 10.1  | 2529   | 29 | AY421056 |    | AY421056 Homo sapi |             |